

---

## Praktikumsaufgaben 4

### WiMa-Praktikum

---

**Aufgabe 1** In einer Studie von Thornes und Collard (1979) wurden Scheidungshintergründe untersucht. Etwa je 500 geschiedene und verheiratete Personen wurden gefragt, ob sie vor ihrer Ehe Geschlechtsverkehr (veg) und ob sie während ihrer Ehe Geschlechtsverkehr (aeg) mit einem anderen Partner außer ihrem (späterem) Ehepartner hatten. Welcher Typ Daten wurde in der Studie erhoben? Führe folgenden Data-Step durch, der zwar recht ungewöhnlich ist, aber gerade bei Feldversuchen die Datenerfassung erheblich verkürzt.

```
DATA divorce;  
DO fstand='gesch', 'verh';  
DO gschl='m', 'w';  
DO veg='ja', 'nein';  
DO aeg='ja', 'nein';  
INPUT anzahl @@; OUTPUT;  
END; END; END; END;  
CARDS;  
28 60 17 68 17 54 36 214  
11 42 4 130 4 25 4 322  
;  
RUN;
```

Untersuchen Sie mittels der Prozedur CATMOD, ob sich die Scheidungswahrscheinlichkeit in dieser Stichprobe auf das Geschlecht, voreheliche und außereheliche Sexualerfahrung zurückführen lässt. Für die Auswertung wurde das log-lineare Modell mit Effekt-Kodierung und einem Interaktionseffekt von vorehelichem und außerehelichem Geschlechtsverkehr gewählt. Für die Effekt-Kodierung von  $X^A$  für  $r$  Ausprägungen gilt:

$$X_j^A = \begin{cases} 1, & \text{falls die } j\text{-te Kategorie von } A \text{ vorliegt} \\ -1, & \text{falls die Kategorie } r \text{ von } A \text{ vorliegt} \\ 0, & \text{sonst.} \end{cases}$$

Des Weiteren soll der gewichtete Kleinste-Quadrate-Schätzer  $\hat{\alpha}$  in Betracht gezogen werden (WLS). In welcher Spalte von welcher Tabelle können die Werte von  $\hat{\alpha}$  in der Ausgabe von PROC CATMOD gefunden werden? Welche Kovariablen könnten bei der Analyse eine Rolle spielen?

Folgender Code kann verwendet werden:

```
PROC CALMOD DATA=divorce ;
    ...;
    RESPONSE ...;
    MODEL ...=... |... / WLS NODESIGN PRED=PROB;
RUN;
```

**Aufgabe 2** In einer Studie wurden Kinder befragt, wie lange sie am Tag Videospiele spielen, wie lange sie schlafen und wie schnell sie 400 Meter schwimmen können. Anschließend wurden die gesammelten Daten in 10 Gruppen zusammengefasst von 30 bis 300 Minuten Spielzeit.

Der Datensatz `kind.dat` enthält die mittleren Schlafzeiten in Stunden und die mittleren Zeiten in Minuten, die ein Kind für 400 Meter schwimmen benötigt. Überprüfen Sie mittels linearer Regression, ob die Dauer des Videospielens einen Einfluss auf die Schlafzeit beziehungsweise auf die Schwimmzeit hat. Verwenden Sie hierfür die Prozedur `REG`. Vergleichen Sie Ihre Ergebnisse anhand der beiden Grafiken, Schätzern und Bestimmtheitsmaßen.

Tipp: Für die Grafiken muss `ODS GRAPHICS` aktiviert werden.

**Aufgabe 3** Bei der Erstellung von Mietspiegeln einer Stadt wird aus der Gesamtheit aller in Frage kommenden Wohnungen eine repräsentative Zufallsstichprobe gezogen (im Fall der Stadt München durch Infratest), und die interessierenden Daten werden von Interviewern anhand von Fragebögen ermittelt. Der vorliegende Datensatz `miete03.asc` stellt einen Ausschnitt aus dem Mietspiegel München des Jahres 2003 dar und enthält die Daten von 2053 Wohnungen. Die Daten sind wie folgt bezeichnet:

| Bez.   | Bedeutung                     | Bez.     | Bedeutung                               |
|--------|-------------------------------|----------|---|
| nm     | Nettomiete in €               | wohngut  | Gute Wohnlage (ja=1, nein=0)            |
| nmqm   | Nettomiete pro $m^2$ in €     | wohnbest | Beste Wohnlage (ja=1, nein=0)           |
| wf1    | Wohnfläche in $m^2$           | ww0      | Warmwasserversorgung (ja=0, nein=1)     |
| rooms  | Anzahl der Zimmer             | badkach0 | Gekacheltes Badezimmer (ja=0, nein=1)   |
| bj     | Baujahr der Wohnung           | zh0      | Zentralheizung (ja=0, nein=1)           |
| bez    | Stadtbezirk                   | badextra | Besondere Badausstattung (ja=1, nein=0) |
| kueche | Gehobene Küche (ja=1, nein=0) |          |   |

Als Modell soll

$$nm = \alpha_0 + \alpha_1 nmqm + \alpha_2 wf1 + \alpha_3 rooms + \alpha_4 bj + \alpha_5 wohngut + \alpha_6 wohnbest + \alpha_7 ww0 + \alpha_8 zh0 + \alpha_9 badkach0 + \alpha_{10} badextra + \alpha_{11} kueche$$

betrachtet werden. Führen Sie eine Regressionsanalyse durch. Testen Sie dabei auch, ob die Variablen `wohngut` und `zh0` überhaupt relevant sind, einzeln und gemeinsam.